Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

Network Theory
Lecture 4.03

EEU45C09 / EEP55C09
Self Organising Technological Networks

Nicola Marchetti
nicola.marchetti@tcd.ie

The World Wide Web

The World Wide Web is the largest network for which topological information is currently available.

The nodes are webpages and the edges are hyperlinks (URLs).

At the end of 1999, this network was close to one billion nodes.

The degree distribution of web pages has been found to follow a power-law over several orders of magnitude.

World Wide Web

Edges in the WWW are directed, so the network has two degree distributions: for incoming and outgoing edges.

The distribution of outgoing edges $P_{\mathrm{out}}(k)$ is the probability that a webpage has k outgoing hyperlinks.

The distribution of incoming edges $P_{\mathrm{in}}(k)$ is the probability that k hyperlinks point to a webpage.

World Wide Web

Several studies have shown that both $P_{\mathrm{out}}(k)$ and $P_{\mathrm{in}}(k)$ have power-law tails:

$$P_{\mathrm{out}}(k) \sim k^{-\gamma_{out}}$$
$$P_{\mathrm{in}}(k) \sim k^{-\gamma_{in}}$$

Over many different studies, there seems to be a characteristic range of γ values centred around:

$$\gamma_{\mathrm{out}} \approx 2.5$$
$$\gamma_{\mathrm{in}} \approx 2.1$$

Interestingly, it seems that γ_{in} seems to stay the same as the web grows. However, γ_{out} tends to grow with both sample size and time.

World Wide Web

Despite its size, the WWW displays the small-world property. In 1999, Albert, Jeong and Barabàsi found $\langle l \rangle \approx 11.2$ for a sample of 325,729 WWW nodes.

They further predicted that a "full" WWW of 800 million nodes would have $\langle l \rangle \approx 19$.

A year later, Broder et al found the average path length in a 50 million node sample of the WWW to be $\langle l \rangle \approx 16$, which agrees well with AJB.

World Wide Web

Since the WWW is directed, we can't immediately calculate the clustering coefficient C. However, treating each directed edge as being undirected, it is possible to proceed.

In 1999, Adamic measured the clustering coefficient C using 50 million web pages. After screening the data somewhat (making the graph undirected, removing some nodes with degree one), she found $C = 0.1078$.

This is orders of magnitude higher than $C_{\mathrm{random}} = 0.00023$ corresponding to a random graph of the same size and average degree.

World Wide Web

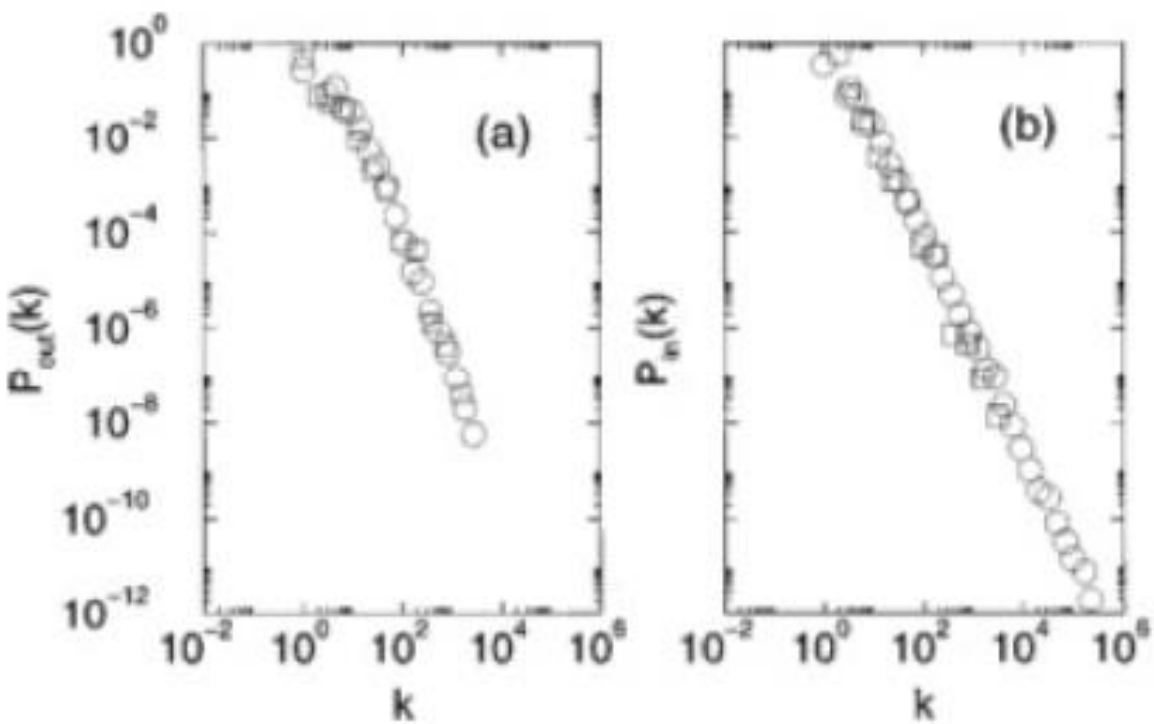


Figure 3: The degree distribution of outgoing and incoming edges of the WWW. The squares are from the 325,729 node sample by Albert et al. The circles are due to measurements over 200 million nodes by Broder.

Real World Networks Vs. Models

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}	ℓ_{real}	ℓ_{rand}	ℓ_{pow}	Reference	Nr.
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999	1
WWW	4×10^7	7		2.38	2.1				Kumar *et al.*, 1999	2
WWW	2×10^8	7.5	4000	2.72	2.1	16	8.85	7.61	Broder *et al.*, 2000	3
WWW, site	260 000				1.94				Huberman and Adamic, 2000	4
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2	4	6.3	5.2	Faloutsos, 1999	5
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999	6
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000	7
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999	8
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b	9
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási *et al.*, 2001	10
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási *et al.*, 2001	11
Sexual contacts*	2810			3.4	3.4				Liljeros *et al.*, 2001	12
Metabolic, *E. coli*	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong *et al.*, 2000	13
Protein, *S. cerev.**	1870	2.39		2.4	2.4				Jeong, Mason, *et al.*, 2001	14
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000	14
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000	16
Citation	783 339	8.57			3				Redner, 1998	17
Phone call	53×10^6	3.16		2.1	2.1				Aiello *et al.*, 2000	18
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001	19
Words, synonyms*	22 311	13.48		2.8	2.8				Yook *et al.*, 2001b	20

Figure 4: Table shows the scaling exponents for the degree distributions of real-world networks. For directed networks, the in-degree and out-degree are listed separately. The average path length is compared for real-networks against predictions for random graphs and WS graphs.

More Power-Law degree distributions

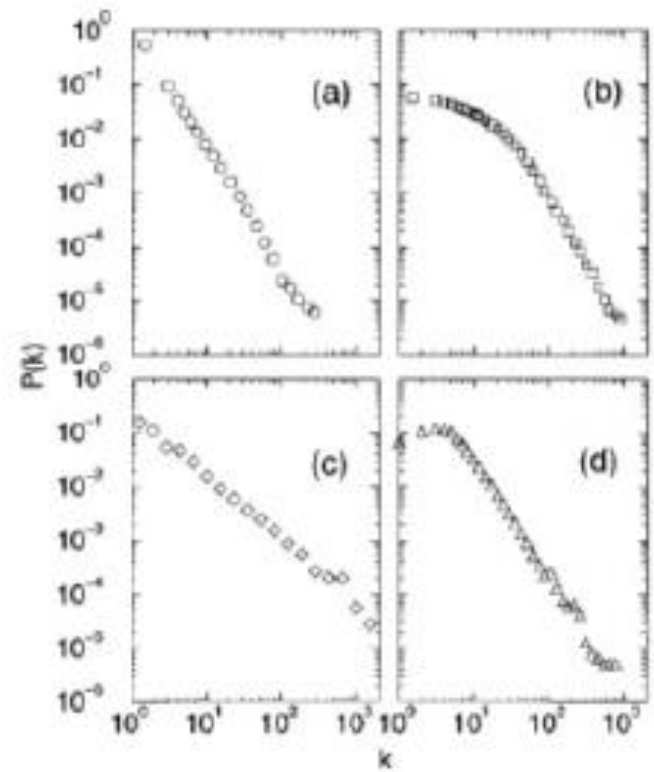


Figure 5: The degree distribution of a) the internet at router level, b) movie actor collaboration, c) co-authorship network of high-energy physicists, d) co-authorship network of neuroscientists.

Scale-Free Networks

We have seen that some real-world networks differ from random graphs in that their degree distribution follows a power-law,

$$P(k) \sim k^{-\gamma}.$$

Since power-laws are independent of scale (look similar at any scale), these networks are called scale-free networks.

Scale-Free Networks

What is the mechanism responsible for the emergence of scale-free networks?

We must shift our focus from modelling some network topology to trying to model the network assembly and evolution.

Scale-Free Networks: Barabási-Albert Model

The origin of the power-law degree distribution observed in networks was first addressed by Barabási and Albert in 1999.

They argued that the scale-free nature of real networks is rooted in two generic mechanisms:

- Growth
- Preferential attachment

Up to now (random graph and small-world network), we start with a fixed number of nodes N and randomly connect or rewire them.

Scale-Free Networks: Barabási-Albert Model

Most real-world networks are open systems which grow by continuously adding nodes eg. internet, WWW, collaboration networks.

Models discussed so far assume that the probability that two nodes connect is independent of their degree. However, in the real world, a web page is more likely to include a link to an already popular site eg. Google.

Therefore, we find that "rich" nodes get "richer". There must be a preferential attachment mechanism.

Using these elements (growth and preferential attachment), the Barabási-Albert model was the first to generate a network with power-law degree distribution.

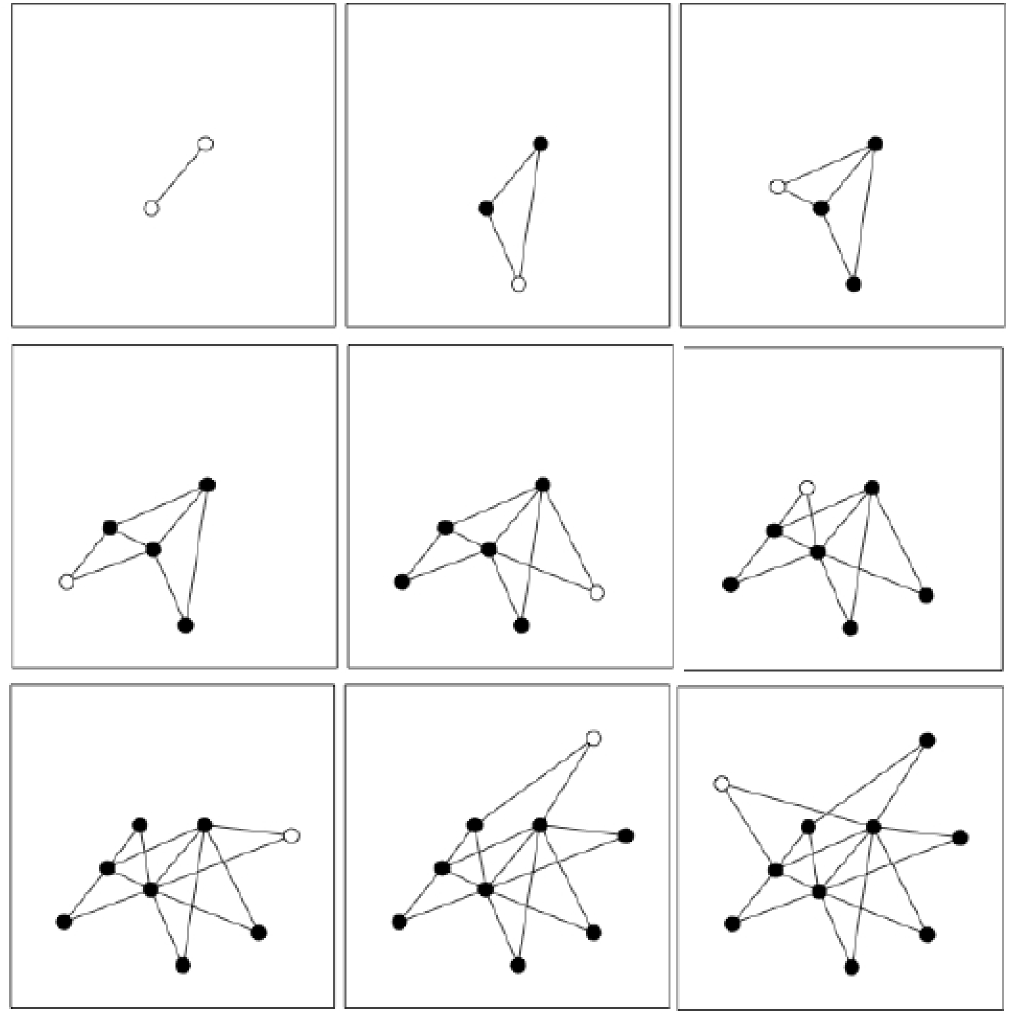
Scale-Free Network: Barabási-Albert Model

This model constructs a scale-free graph that has a power-law degree distribution. The model used preferential attachment to create nodes with many links:

1. We start with m_0 nodes, the links between which are chosen arbitrarily, as long as each node has at least one link.
2. At every time step, attach a new node to this core with $m \leq m_0$ edges. The probability of attaching to an existing node i, depends on its degree:

$$p_i(k_i) = \frac{k_i}{\sum_j k_j}. \qquad \textbf{(1)}$$

3. Repeat Step 2 until you have the appropriate sized graph.



The sequence of images shows nine subsequent steps of the Barabási-Albert model. Empty circles mark the newly added node to the network, which decides where to connect its two links (m=2) using preferential attachment **(1)**

Scale-Free Network: Barabási-Albert Model

After time t, this procedure results in a network with $N = t + m_0$ nodes and mt new edges.

BA model networks are characterised by a degree distribution which follows a power law,

$$P(k) \sim k^{-\gamma},$$

where $\gamma = 3$.

Scale-free networks occur in systems where attachment of new nodes to already popular nodes is very likely.

Scale-Free Network: Barabási-Albert Model

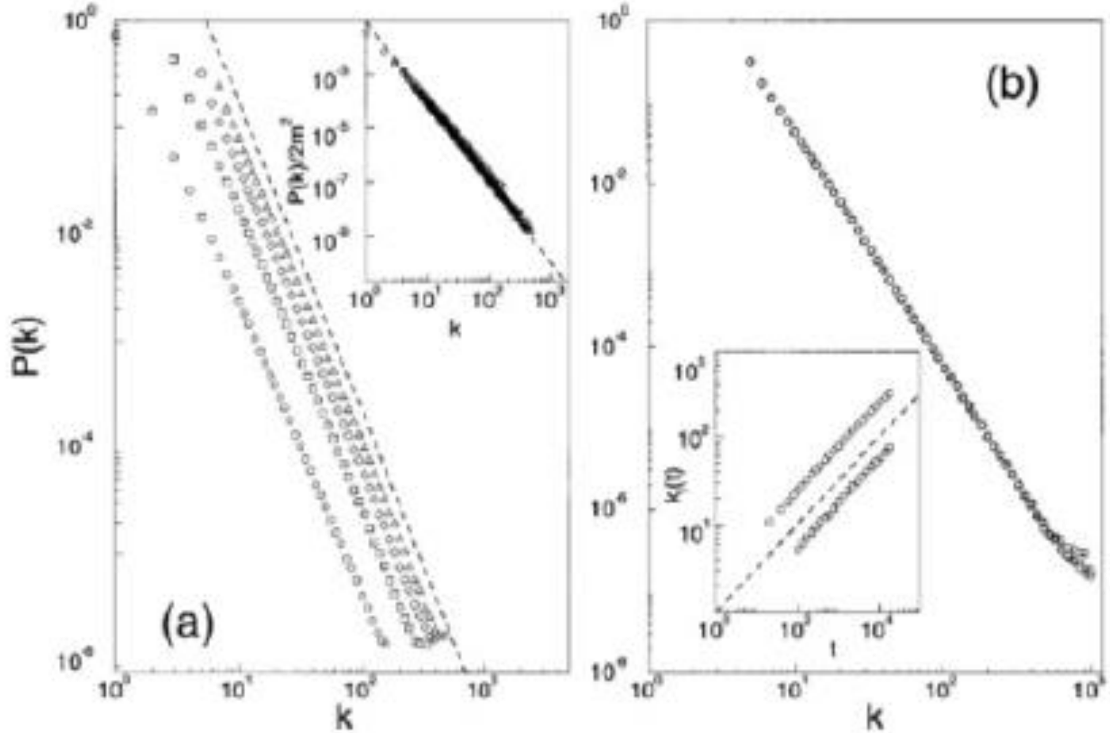


FIG. 21. Numerical simulations of network evolution: (a) Degree distribution of the Barabási-Albert model, with $N=m_0+t=300\,000$ and ○, $m_0=m=1$; □, $m_0=m=3$; ◇, $m_0=m=5$; and △, $m_0=m=7$. The slope of the dashed line is $\gamma=2.9$, providing the best fit to the data. The inset shows the rescaled distribution (see text) $P(k)/2m^2$ for the same values of m, the slope of the dashed line being $\gamma=3$; (b) $P(k)$ for $m_0=m=5$ and various system sizes, ○, $N=100\,000$; □, $N=150\,000$; ◇, $N=200\,000$. The inset shows the time evolution for the degree of two vertices, added to the system at $t_1=5$ and $t_2=95$. Here $m_0=m=5$, and the dashed line has slope 0.5, as predicted by Eq. (81). After Barabási, Albert, and Jeong (1999).

BA model: A limiting case

How do we know that both growth and preferential attachment are necessary to achieve a power-law degree distribution?

Define two limiting cases of the BA model:

- Model A (growth only): Start with a small number of nodes (m_0) and at each step add a new node with $m \leq m_0$ edges. Attach the new node to older nodes with equal probability $p_i(k_i) = 1/(m_0 + t - 1)$.
- Model B (preferential attachment only): Start with N nodes and no edges. At each timestep a node is randomly selected and attached to another node of degree k with probability $p_i(k_i) = k_i / \sum_j k_j$.

nodes with k=0 are assumed to have k=1, otherwise they can not acquire links.

BA model: A limiting case

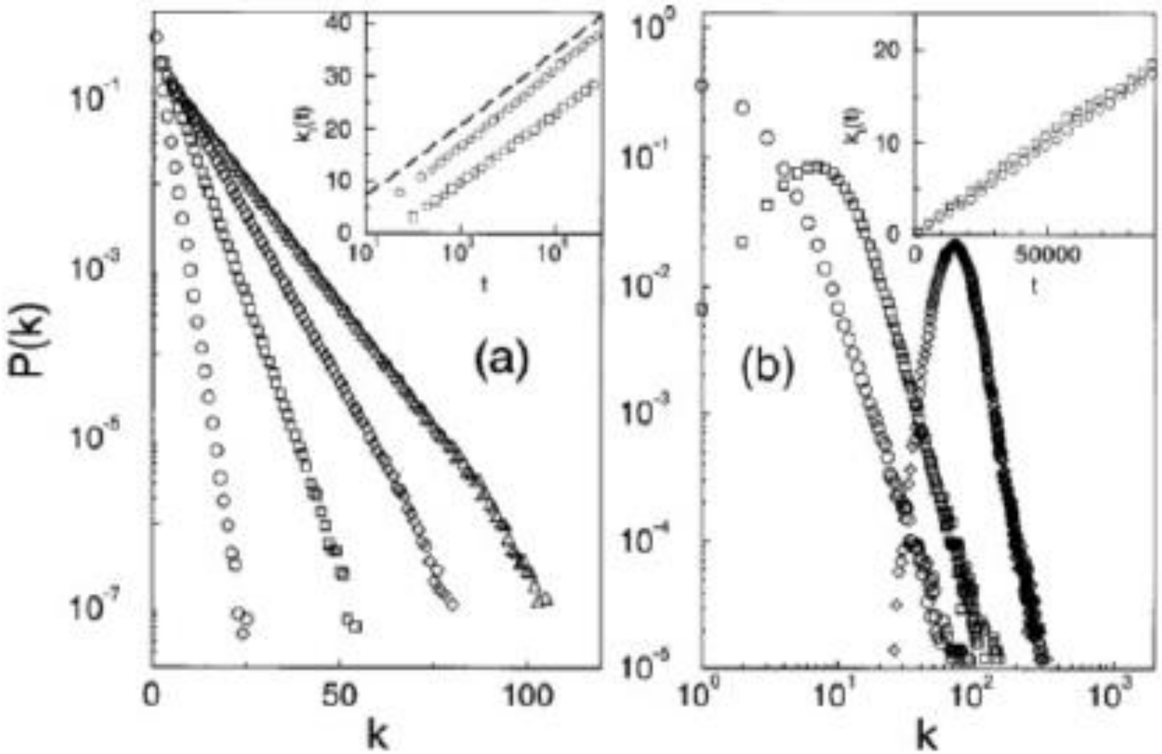


FIG. 22. Degree distribution for two models: (a) Degree distribution for model A: ○, $m_0 = m = 1$; □, $m_0 = m = 3$; ◇, $m_0 = m = 5$; △, $m_0 = m = 7$. The size of the network is $N = 800\,000$. Inset: time evolution for the degree of two vertices added to the system at $t_1 = 7$ and $t_2 = 97$. Here $m_0 = m = 3$. The dashed line follows $k_i(t) = m \ln(m_0 + t - 1)$; (b) the degree distribution for model B for $N = 10\,000$: ○, $t = N$; □, $t = 5N$; and ◇, $t = 40N$. Inset: time dependence of the degrees of two vertices. The system size is $N = 10\,000$. After Barabási, Albert, and Jeong (1999).

Scale-Free Network: Barabási-Albert Model

What do BA model networks look like?

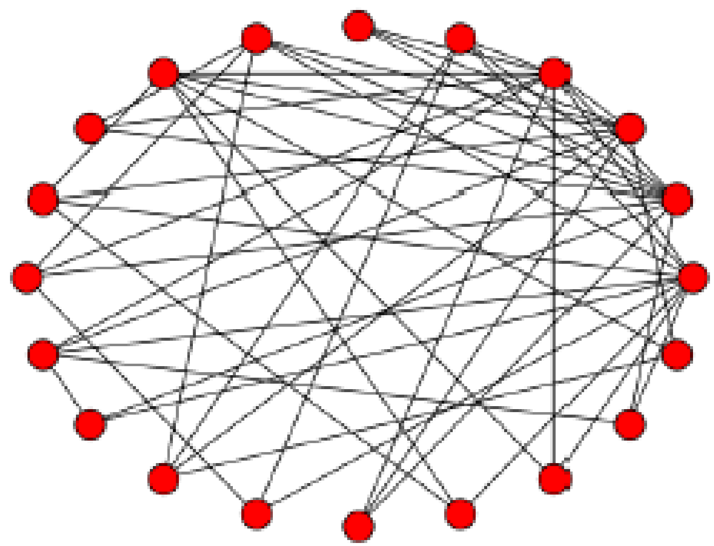


Figure 6: Barabási-Albert model [1]: $N = 20$, $m = 3$

Can you blame anyone who thinks this is a random graph?

A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

M. Newman, *Networks: An Introduction.* **New York, NY, USA: Oxford University Press, Inc., 2010.**

R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, pp. 47–97, Jan 2002.

Ackowledgement

- Dr Neal McBride, Arista Networks